

QWIP: A Quantitative Metric for Quality Control of Aquatic Reflectance Spectral Shape using the Apparent Visible Wavelength

Heidi M. Dierssen^{1*}, Ryan A. Vandermeulen², Brian B. Barnes³, Alexandre Castagna⁴, Els Knaeps⁵, Quinten Vanhellemont⁶

¹University of Connecticut, United States, ²National Aeronautics and Space Administration (NASA), United States, ³University of South Florida, United States, ⁴Ghent University, Belgium, ⁵Flemish Institute for Technological Research (VITO), Belgium, ⁶Royal Belgian Institute of Natural Sciences, Belgium

Submitted to Journal:
Frontiers in Remote Sensing

Specialty Section:
Multi- and Hyper-Spectral Imaging

Article type:
Original Research Article

Manuscript ID:
869611

Received on:
04 Feb 2022

Revised on:
15 Mar 2022

Journal website link:
www.frontiersin.org

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

H.D. and R.V. wrote and conceived of this approach and did the primary data analysis and manuscript preparation. B.B. calculated Wei scores. B.B., A.C., E.K. and Q.V. contributed data, wrote methods, and contributed to the final editing of the manuscript.

Keywords

remote sensing reflectance, Ocean Color, hyperspectral remote sensing, Hydrologic optics, Water Quality, QA/QC - Quality Assurance / Quality Control, Water-leaving reflectance spectra

Abstract

Word count: 281

The colors of the ocean and inland waters span clear blue to turbid brown, and the corresponding spectral shapes of the water-leaving signal are diverse depending on the various types and concentrations of phytoplankton, sediment, detritus and colored dissolved organic matter. Here we present a simple metric developed from a global dataset spanning blue, green and brown water types to assess the quality of a measured or derived aquatic spectrum. The Quality Water Index Polynomial (QWIP) is founded on the Apparent Visible Wavelength (AVW), a one-dimensional geophysical metric of color that is inherently correlated to spectral shape calculated as a weighted harmonic mean across visible wavelengths. The QWIP represents a polynomial relationship between the hyperspectral AVW and a Normalized Difference Index (NDI) using red and green wavelengths. The QWIP score represents the difference between a spectrum's AVW and NDI and the QWIP polynomial. The approach is tested extensively with both raw and quality controlled field data to identify spectra that fall outside the general trends observed in aquatic optics. For example, QWIP scores less than or greater than 0.2 would fail an initial screening and be subject to additional quality control. Common outliers tend to have spectral features related to: 1) incorrect removal of surface reflected skylight or 2) optically shallow water. The approach was applied to hyperspectral imagery from the Hyperspectral Imager for the Coastal Ocean (HICO), as well as to multispectral imagery from the Visual Infrared Imaging Radiometer Suite (VIIRS) using sensor-specific extrapolations to approximate AVW. This simple approach can be rapidly implemented in ocean color processing chains to provide a level of uncertainty about a measured or retrieved spectrum and flag questionable or unusual spectra for further analysis.

Contribution to the field

The QWIP approach provides a simple quantitative tool to evaluate the quality of both field and satellite water spectra. Because of its simplicity and highly visual component, we foresee high applicability for the QWIP method across a wide array of applications. The quality of various atmospheric correction routines can be quickly assessed by comparing retrieved results to the QWIP. The QWIP can also be used to fine tune approaches used to remove surface reflected skylight and other data processing choices from above water reflectance measurements. The QWIP method provides a simple tool to help evaluate spectral shape and magnitude for a variety of aquatic water-leaving reflectance spectra.

Funding statement

Funding for this research was provided by BELSPO Stereo III TIMBERS project (SR/00/381) and NASA Ocean Biology and Biogeochemistry (80NSSC20K1518). A.C. acknowledges funding provided by BELSPO Stereo III PHYTOBEL project (SR/02/213).

Ethics statements

Studies involving animal subjects

Generated Statement: No animal studies are presented in this manuscript.

Studies involving human subjects

Generated Statement: No human studies are presented in this manuscript.

Inclusion of identifiable human data

Generated Statement: No potentially identifiable human images or data is presented in this study.

Data availability statement

Generated Statement: The data analyzed in this study is subject to the following licenses/restrictions: The CASCK data used to develop the QWIP is provided in Supplemental Materials. A data paper describing the PANTHYR data is in preparation. One dataset was a raw WISP dataset including bad data for purposes of quality analysis that will be published in its final form in a data paper in preparation.. Requests to access these datasets should be directed to heidi.dierssen@uconn.edu.

In review

QWIP: A Quantitative Metric for Quality Control of Aquatic Reflectance Spectral Shape using the Apparent Visible Wavelength

1 Heidi M. Dierssen¹, Ryan A. Vandermeulen^{2,3}, Brian B. Barnes⁴, Alexandre Castagna⁵, Els
2 Knaeps⁶, and Quinten Vanhellemont⁷

3 ¹Department of Marine Sciences, University of Connecticut, Groton, CT, USA

4 ²Ocean Ecology Laboratory, Goddard Space Flight Center, National Aeronautics and Space
5 Administration, Greenbelt, MD, USA.

6 ³Science Systems and Applications Inc., Lanham, MD, USA

7 ⁴College of Marine Science, University of South Florida, St. Petersburg, FL, USA

8 ⁵Protistology and Aquatic Ecology, Department of Biology, Ghent University, Ghent, Belgium

9 ⁶Flemish Institute for Technological Research (VITO), Mol, Belgium

10 ⁷Royal Belgian Institute of Natural Sciences, Brussels, Belgium

11

12 * **Correspondence:**

13 Corresponding Author

14 heidi.dierssen@uconn.edu

15 **Keywords:** remote sensing reflectance, ocean color, hyperspectral remote sensing, hydrologic
16 optics

17 Abstract

18 The colors of the ocean and inland waters span clear blue to turbid brown, and the corresponding
19 spectral shapes of the water-leaving signal are diverse depending on the various types and
20 concentrations of phytoplankton, sediment, detritus and colored dissolved organic matter. Here we
21 present a simple metric developed from a global dataset spanning blue, green and brown water types
22 to assess the quality of a measured or derived aquatic spectrum. The Quality Water Index
23 Polynomial (QWIP) is founded on the Apparent Visible Wavelength (AVW), a one-dimensional
24 geophysical metric of color that is inherently correlated to spectral shape calculated as a weighted
25 harmonic mean across visible wavelengths. The QWIP represents a polynomial relationship between
26 the hyperspectral AVW and a Normalized Difference Index (NDI) using red and green wavelengths.
27 The QWIP score represents the difference between a spectrum's AVW and NDI and the QWIP
28 polynomial. The approach is tested extensively with both raw and quality controlled field data to
29 identify spectra that fall outside the general trends observed in aquatic optics. For example, QWIP
30 scores less than or greater than 0.2 would fail an initial screening and be subject to additional quality
31 control. Common outliers tend to have spectral features related to: 1) incorrect removal of surface
32 reflected skylight or 2) optically shallow water. The approach was applied to hyperspectral imagery
33 from the Hyperspectral Imager for the Coastal Ocean (HICO), as well as to multispectral imagery
34 from the Visual Infrared Imaging Radiometer Suite (VIIRS) using sensor-specific extrapolations to
35 approximate AVW. This simple approach can be rapidly implemented in ocean color processing

36 chains to provide a level of uncertainty about a measured or retrieved spectrum and flag questionable
37 or unusual spectra for further analysis.

38 **1 Introduction**

39 The color of a water body is a complex mixture of light that has been reflected from the water surface
40 (sky and other floating substances) and reflected from within the water column (water color). The
41 aquatic optics community refers to the water color reflectance as “Water-leaving Radiance
42 Reflectance” (R_w) or more commonly, but less descriptively, “Remote Sensing Reflectance” (R_{rs}).
43 Many methods are available to estimate R_{rs} both from instruments within and above the water surface
44 and from aircraft and satellites orbiting the Earth. Each method to approximate R_{rs} involves some
45 level of estimation to either remove the specular reflectance from the air-water interface form the
46 above water measurement or to propagate the underwater upwelling signal to and across the
47 interface, as well as to compensate for instrument shading and other measurement artifacts (Ruddick
48 et al., 2019; Zibordi et al., 2019; Lee et al., 2020). To meet the needs of a growing list of applications
49 for water color imagery (Dierssen et al., 2021), more automated data analysis is essential to develop
50 and validate different approaches to assess regional and global aquatic optical properties and
51 phytoplankton and benthic biodiversity. Automated systems are being developed to estimate R_{rs} from
52 drones, moorings, profilers, and offshore platforms (summarized in Dierssen et al., 2020). As
53 systems become more automated going forward and more “big data” hyperspectral datasets are
54 generated routinely for use by the broader science community, new metrics are needed to provide
55 automated quality control of data from different water types, as well as to assess the quality of
56 satellite-retrievals of R_{rs} .

57 **How can I assess the data quality of a water spectrum?** This simple sounding question is quite
58 complex to answer objectively because R_{rs} is a derived parameter and the uncertainties are quite
59 challenging to quantify for each method under a wide array of environmental conditions. Spectra are
60 not only influenced by the optical properties of a large diversity of dissolved and particulate
61 components and the potential contribution from the benthos, but also by measurement artifacts
62 related to the solar and viewing angles, sky and wave conditions, air-water interface, platform
63 disturbance, spatial inhomogeneity in the water column, and distance from land masses (Voss et al.,
64 2017; Bulgarelli and Zibordi, 2020; Shang et al., 2020). If we screen out field data to encompass only
65 a set of predefined or “ideal” conditions with calm seas, clear skies, and constrained solar and
66 viewing angles, for example, then the measurement uncertainty is decreased significantly. This can
67 be justified for the generation of fiducial reference spectra for satellite valiation and calibration, such
68 as data from the Aeronet-OC program (Zibordi et al., 2009). However, as noted by a recent
69 intercomparison exercise, there can still be considerable uncertainty in R_{rs} even under ideal
70 conditions and with well-calibrated sensors (Tilstone et al., 2020). Moreover, such ideal conditions
71 are rarely met in the field and represent only a fraction of the diverse environmental conditions
72 encountered in natural ecosystems across latitudes and seasons. Retrieval models applied to satellite
73 imagery also contain a much broader set of conditions, in terms of water column, atmosphere and
74 surface states, as well as observation and illumination geometries. As retrieval approaches are
75 diversifying, the community is finding that even non-optimal spectral information can still be useful
76 for algorithm development and parameter validation for a variety of applications. For example,
77 algorithms with narrowband indices or relying on derivatives may be less influenced by surface
78 reflected sunlight than full inversion-type algorithms (Dierssen et al., 2021).

79 Many standard options are available for processing radiometric data, including flags for low solar
80 zenith angles and low light conditions and for identifying rain, outlined in processing software such

81 as HyperInSpace (Aurin, 2022). Operationally, site-specific criteria and thresholds are often
82 developed based on careful examination of the data from that region. For example, spurious ship-
83 based radiometric data were identified using five different spectral shape metrics tuned for highly
84 absorbing and weakly scattering conditions characteristic of the Baltic Sea (Qin et al., 2017). The
85 thresholds in such metrics are set based on human evaluation and interpretation of the expected data
86 for a region.

87 **Does my water spectrum look like other spectra?** For many aquatic data, there are clear standards
88 that can be used to assess the quality of data. Such metrics are still being defined in the aquatic optics
89 community due to the large range in environmental conditions and the many corrections that must be
90 considered for any measurement (e.g., most absorption measurements have an associated scattering
91 correction and most scattering measurements have an associated absorption correction). Commonly,
92 intercomparisons between instruments and methods are the primary means to assess the operational
93 uncertainties (Tilstone et al., 2020) or through simulation and closure studies (Zaneveld, 1994;
94 Tzortziou et al., 2006), although assessing the accuracy of a measurement still remains an
95 outstanding problem (Ruddick et al., 2019; Zibordi et al., 2019). In this paper we do not assess the
96 accuracy of R_{rs} , but rather develop a method to assess the quality of a water spectrum using data
97 collected and filtered using community standard practices.

98 Experienced researchers tend to have a “gut” sense for what looks like a reasonable spectrum, but
99 definitive approaches for assessing spectral shape are limited. Questionable spectral shapes are
100 readily identifiable when reflectance values in the near-infrared wavelengths are much higher than
101 anticipated for the type of water collected or when spectra exhibit an exponentially increasing tail
102 from blue to ultraviolet, indicating that reflected diffuse skylight may not be removed fully from the
103 spectrum (Mobley, 1999; Gould et al., 2001). A recent work by Wei et al. (2016) provides a means of
104 scoring a spectrum compared to 23 different simulated optical water types at a maximum of 9
105 wavebands (Wei et al., 2016). This spectral matching approach provides a score as to how close a
106 spectrum matches one of the predefined water types. As noted below, this approach may be limited to
107 standard simulated conditions and may not represent all water types encountered across the vast
108 aquascape (Barnes et al., 2019).

109 **Are there other simple metrics that can be used to assess the quality of a water spectrum?** Here,
110 we present a simple quantitative index approach to conduct quality control of an R_{rs} spectrum and
111 thereby determine whether it appears similar to other ocean color spectra from a wide variety of blue,
112 green and brown water types. The **Quality Water Index Polynomial (QWIP)** is founded on the
113 Apparent Visible Wavelength (AVW) a one-dimensional geophysical metric of “color”, calculated as
114 the weighted harmonic mean of the reflectance spectrum across a range of wavelengths
115 (Vandermeulen et al., 2020). This metric reduces the hyperspectral or multispectral (after applying
116 sensor-specific correction factors) data to a continuous variable representing the mean “color” in
117 wavelength (expressed in nm). As noted in Figure 10b of Vandermeulen et al. (2020), a good
118 correlation is found between AVW and chlorophyll *a* concentrations using a global dataset. Building
119 upon this, we developed a relationship between AVW and standard multi-channel waveband indices
120 to identify spectra that fall outside the general trends observed in aquatic optics for optically deep
121 waters. The approach was developed with a large global dataset representing blue, green, and brown
122 waters and was further tested extensively with field and satellite datasets. This simple approach can
123 be rapidly implemented in ocean color processing chains to provide a level of uncertainty about the
124 spectrum and flag questionable or unusual spectra for further analysis.

125 **2 Methods**

126 The method was developed using a large global dataset of remote sensing reflectance compiled from
 127 different studies (CASCK-P dataset) and then tested using several different regional field datasets
 128 collected with above-water methodology and on satellite-retrievals of water-leaving reflectance data.

129 **2.1 Field Datasets**

130 **CASCK dataset (Casey, Castagna and Knaeps).** The majority of the training dataset was obtained
 131 from a recent global compilation of hyperspectral optical data that includes profiled and buoy
 132 mounted in-water radiometers, as well as ship-mounted and hand-held above-water methods (Casey
 133 et al., 2020). Because intense green and brown water spectra are underrepresented in this dataset, we
 134 have augmented it with two additional coastal and inland water datasets. As such, a dataset
 135 comprising primarily of green water spectra with pronounced red edge reflectance, collected with a
 136 hand-held single spectrometer using the skylight-blocked method in nine different Belgian lakes and
 137 a lagoon was included (Castagna et al., 2020). Above-water spectra of brown water from the Scheldt
 138 River Delta were also included (Castagna et al., submitted). An additional dataset collected with a
 139 hand-held single radiometer system in the river deltas of the Scheldt (Belgium), Gironde (France),
 140 and Río de la Plata (Argentina) with total suspended matter up to 1400 g m^{-3} was also incorporated
 141 to represent extremely turbid conditions (Knaeps et al., 2015). In total, the CASCK dataset contains
 142 1029 R_{rs} spectra, and is provided in the Supplementary Materials. The global distribution of the
 143 CASCK dataset is shown in Figure 1.

144 **PANTHYR Venice and Ostend datasets.** Data from two autonomous hyperspectral PANTHYR
 145 radiometer systems (Vansteenvagen et al., 2019) deployed in the Adriatic Sea at the Acqua Alta
 146 Oceanographic Tower in the Gulf of Venice, Italy (AAOT, 45.3139°N , 12.5083°E) ($n=1622$) and in
 147 the North Sea at Research Tower 1 near Ostend, Belgium (RT1, 51.2464°N , 2.9193°E) ($n=4945$)
 148 from September to December 2019 were also included (Vanhellemont, 2020). The locations of the
 149 two towers is shown in Figure 1. These data were subject to considerable quality control and
 150 assessment including removing spectra collected under sub-optimal conditions as described in
 151 Vanhellemont (2020). A data paper further describing these data in final format is forthcoming.

152 **WISP-3 Raw dataset.** Above-water data have been collected using the hand-held WISP-3
 153 measurements as part of the Belgian coast Lifewatch program since April 2019, (Mortelmans et al.,
 154 2019). For this analysis, no quality assessment criteria were applied and the dataset included all
 155 training measurements including spectra taken of land, on deck, and with lens cap on in order to test
 156 the QWIP on wide array of good and bad spectra ($n=869$). The instrument has three hyperspectral
 157 radiometers that simultaneously capture the downwelling plane irradiance, E_d , upwelling water
 158 system (surface + water-leaving) radiance (L_{ws}), and skylight radiance (L_{sky}) (Hommersom et al.,
 159 2012) The data were processed to remote sensing reflectance R_{rs} using a solar-zenith angle
 160 dependent ρ factor to account for spectral sea surface reflectance of skylight (Zhang et al., 2017).
 161 The ρ factor was calculated for a viewing angle of 40° to water, 135° azimuth from sun with an
 162 Aerosol Optical Thickness at 555 nm of 0.1. For these data, a wind speed of 5 m s^{-1} was used and
 163 data collected under solar zenith angles greater than 60° , outside the scope of typical values, used a ρ
 164 comparable to SZA of 60° . Mean visible ρ used here ranged from 0.032 to 0.036. Residual skylight
 165 was removed with a baseline subtraction following a semi-analytical correction using two narrow
 166 band features in the near-infrared red at 715 nm and 735 nm following Gould et al. (2001). A
 167 spectrally flat residual baseline correction, B , was estimated from the measurements of radiance

168 reflectance of skylight ($R_{\text{sky}} = L_{\text{sky}}/E_{\text{d}}$) and of the water system including water-leaving and surface
169 reflected signal ($R_{\text{ws}} = L_{\text{ws}}/E_{\text{d}}$) according to:

170

$$B = R_{\text{ws}}(735) - \rho(735) R_{\text{sky}}(735) - \frac{R_{\text{ws}}(715) - R_{\text{ws}}(735)}{\left(\frac{a_{\text{w}}(735)}{a_{\text{w}}(715)} - 1\right)} \quad (1)$$

171 This correction assumes that absorption at 715 nm and 735 nm is dominated by pure water (a_{w}) and
172 that the combined effects of backscattering, b_{b} , and the f/Q bidirectional factor, and residual skylight
173 are well approximated by a spectrally flat value (equivalent in magnitude) between 715 nm and 735
174 nm. It is further assumed that the residual skylight signal is well approximated by B in the visible
175 range. The assumption of spectrally flat pattern of b_{b} and f/Q in this narrow range is well justified
176 (cf. Ruddick et al., 2006). The absorption by pure water at 715 nm and 735 nm were set at 1.0242 m^{-1}
177 and 2.2780 m^{-1} , respectively (Röttgers et al., 2016).

178 **2.2 Satellite Datasets**

179 **HICO dataset.** We tested the proposed radiometric quality control procedure on a series of retrieved
180 scenes from the Hyperspectral Imager for the Coastal Ocean (HICO), in order to examine the
181 algorithm's efficacy in the identification of low-quality satellite returns. HICO Level-1B files were
182 downloaded from the NASA Ocean Biology Processing Group (<https://oceancolor.gsfc.nasa.gov/l2>),
183 and processed using l2gen program packaged as part of the NASA SeaDAS (Ocean Biology
184 Processing Group, 2022). A heritage atmospheric correction procedure was used (Gordon and Wang
185 1994, Bailey et al. 2010), with the additional use of the ATmospheric REMoval code (ATREM; Gao
186 and Davis 1997) built into l2gen, which provides hyperspectral compensation of the water vapor
187 absorption for the atmospheric correction process, following Ibrahim et al. (2018). Data were
188 processed to output a pixel-wise continuous spectra of R_{rs} from 398 nm to 702 nm, including all
189 standard SeaDAS Level-2 flags. This includes the default masking threshold in l2gen that can
190 exclude the most turbid waters. The AVW (Eq. 2) over the range of 400 nm to 700 nm was
191 subsequently calculated for every pixel after interpolation of the spectrum to 1 nm resolution using
192 cubic splines.

193 **VIIRS Matchup Dataset.** In an effort to examine the impact of applying automated quality control
194 criteria on a multi-spectral validation stream, we retrieved all R_{rs} validation matchups for the Visual
195 Infrared Imaging Radiometer Suite (SNPP-VIIRS) from the SeaWiFS Bio-optical Archive and
196 Storage System (SeaBASS; <https://seabass.gsfc.nasa.gov/search#val>), for all SeaBASS and
197 AERONET matchups ($n = 2850$). The exclusion criteria followed NASA processing
198 recommendations (Bailey and Werdell, 2006). An empirical conversion of the multispectral AVW
199 values to a hyperspectral-equivalent AVW was applied to the VIIRS satellite and *in situ* matchups
200 prior to analysis (Vandermeulen et al., 2020; Vandermeulen, 2022).

201 **2.3 Water color classification**

202 A simple decision tree was implemented to determine whether the spectral shape of R_{rs} would be
203 classified as blue-green, green, or brown in color following from the approach presented in
204 (Balasubramanian et al., 2020). This simplified classification (Figure 2) was used to indicate where
205 different types of waters fall within the QWIP schema.

206 **2.4 Wei Score.**

207 Quality of all spectra was assessed according to the Wei et al. (2016) approach, using the Matlab-
 208 based code provided therein. Specifically, a multispectral subset was extracted for each spectrum (at
 209 wavelengths (nm): 412, 443, 488, 510, 531, 547, 555, 667, and 678), which was then normalized and
 210 compared to each of 23 reference spectra representing disparate optical water types. A spectral angle
 211 mapper was used to identify the most spectrally similar reference spectrum. Normalized reflectance
 212 data at each wavelength of the subset were then compared to the corresponding reference spectrum
 213 and associated boundaries. The final QA score was determined as the number wavelengths for which
 214 the reflectance datum fit within the reference boundaries, divided by the total number of wavelengths
 215 assessed (9 for this work). Thus QA scores ranged from 0 (at no wavelength does the target spectrum
 216 fit within boundaries of the identified water type) to 1 (target spectrum is fully within identified water
 217 type boundaries). Wei et al. (2016) qualitatively discusses “very high scores (>0.9)” and “very low
 218 QA scores (<0.5),” and reports that 90 % of the evaluation spectra had “high QA scores” of >0.8.
 219 Based on this, a Wei score of >0.5 was considered as “Passing” and ≤ 0.5 was considered “Failing.”

220 **2.5 Statistical Analyses**

221 The equations used in the QWIP calculation include the Apparent Visible Wavelength (AVW)
 222 calculated from 400 nm to 700 nm at 1 nm intervals and the Normalized Difference Index (NDI) at
 223 two wavelengths as formulated below. Any negative values of R_{rs} are included in the calculations.
 224 We note that the QWIP and NDI acronyms are used in the equations below for clarity, such that:

$$225 \quad AVW = \left(\frac{\sum_{i=1}^n R_{rs}(\lambda_i)}{\sum_{i=1}^n \frac{R_{rs}(\lambda_i)}{\lambda_i}} \right) \quad (2)$$

$$226 \quad NDI = \frac{(R_{rs}(\lambda_2) - R_{rs}(\lambda_1))}{(R_{rs}(\lambda_2) + R_{rs}(\lambda_1))} \quad (3)$$

227

228 Statistical tests were conducted in Matlab (The Mathworks, Inc.). Model II regression analysis was
 229 used presuming similar uncertainty magnitudes in the involved variables.

230 **2.6 QWIP Development**

231 The algorithm is first developed using a global dataset covering a diverse set of (mostly) optically
 232 deep water types. A training dataset was compiled ($n=1629$) that included the CASCK data (see
 233 Supplemental Materials) and 300 random points selected from each of the PANTHYR datasets,
 234 dubbed here the “CASCK-P” dataset (see **Figure 1**). We evaluated the relationships between AVW
 235 (the mean “color” metric, **Eq. 2**) and other indices to find trends across the wide range of spectral
 236 shapes in the dataset. The objective was to find a simple metric using common spectral bands found
 237 in a variety of multi-spectral datasets (i.e., wavebands from the VIIRS sensor wavelengths were
 238 selected here) where the central tendency of data formed a well-described continuum across the wide
 239 range of AVW values and the amount of deviations from the central tendency could be easily scored.
 240 To better visualize the data trends, the water color was further differentiated into the categories of
 241 blue-green (blue dots), green (green dots), and brown (red dots) following from the decision tree (see
 242 **Figure 2**). Blue-green waters have AVW values ranging from 400 nm to 510 nm and green waters

243 have AVW from 510 nm to 590 nm. The AVW for brown waters overlaps with the upper end of the
 244 green waters, ranging from 555 nm to 575 nm.

245 The QWIP is formally a mathematical model relating an optical index to the AVW. Different
 246 examples of the relationships between AVW and other indices are provided in **Fig. 3** for illustrative
 247 purposes. Blue-to-green band ratios, like those used in chlorophyll algorithms and Vandermeulen et
 248 al. (2020), showed a reasonable overall trend, but resulted in the values being spread across a log
 249 scale with clustering of the green and red spectra (e.g., **Figure 3A**). Band difference algorithms
 250 provided lower predictive power, especially for the brown waters which had much higher differences
 251 than the blue and green waters (**Figure 3B**). Relationships between AVW and maximum wavelength
 252 in visible wavelengths (data not shown) were effective for blue and green waters, but proved to have
 253 low predictive power for intense green algal blooms with high red edge values and turbid brown
 254 waters. Brown waters can have variable peak wavelengths from green to red to near infrared
 255 wavelengths (see **Figure 2**).

256 The NDI (**Eq. 3**) provided a means to highlight the variability of logarithmically distributed data on a
 257 linear scale such that the distance either above or below the central tendency was scored with a
 258 positive or negative value. Different combinations of NDI were systematically evaluated using
 259 wavebands found on historic ocean color sensors. For example, the wavebands used in the standard
 260 chlorophyll *a* concentration algorithm (492 nm and 560 nm) (O'Reilly et al., 1998) did not have high
 261 rank correlation with the AVW for green and brown waters (**Figure 3C**). The best relationship was
 262 found using the NDI calculated with blue/green ($\lambda_1 = 492$ nm) and red ($\lambda_2 = 665$ nm) bands (**Figure**
 263 **4A**). A 4th degree polynomial fit between NDI(492,665) and AVW described the variability across
 264 the AVW range ($R^2=0.974$). This QWIP relationship followed the overall objectives in that the
 265 central tendency was clearly outlined across a wide range of data and distance was easily scored. To
 266 implement, the user would calculate their AVW and NDI(492,665) for a spectrum following from
 267 **Equations 2 and 3** and then calculate the QWIP Score as the difference between the measured and
 268 the predicted NDI based on the QWIP.. Note that the acronyms NDI, AVW, and QWIP are used in
 269 the equations below for clarity and the five coefficients in **Eq. 4** correspond to five variables *p*
 270 provided below the equation:

$$271 \quad \text{QWIP} = p_1 \text{AVW}^4 + p_2 \text{AVW}^3 + p_3 \text{AVW}^2 + p_4 \text{AVW} + p_5 \quad (4)$$

$$272 \quad p = \langle -8.399885 \times 10^{-9}, 1.715532 \times 10^{-5}, -1.301670 \times 10^{-2}, 4.357838 \times 10^0, -5.449532 \times 10^2 \rangle$$

$$273 \quad \text{QWIP score} = \text{NDI}(492,665) - \text{QWIP} \quad (5)$$

274 As shown in **Figure 4A**, 97.5 % of the CASCK-P data used in the calibration of the QWIP fell within
 275 ± 0.1 of the QWIP and 99.4 % within ± 0.2 (**Figure 4B**). Outlier data with QWIP scores greater than
 276 ± 0.2 were subject to additional screening to determine any evident spectral anomalies. The points
 277 with higher negative QWIP scores (**Figure 4C**) appeared to be related to some types of optically
 278 shallow water where the red wavelengths were much lower than anticipated for highly green peaked
 279 waters due to water absorption between the scattering seafloor and the sea surface (e.g., Dierssen et
 280 al., 2003) suggesting the QWIP score may prove a useful metric for identifying certain types of
 281 optically shallow water. Data with high positive QWIP scores often had rising tails in the blue end of
 282 the spectrum consistent with data containing residual surface reflected skylight (Gould et al., 2001)
 283 (**Figure 4C**). In such cases, the AVW would be weighted incorrectly towards the blue end of the
 284 spectrum resulting in a positive QWIP score.

285

286 **3 Results**287 The algorithm is evaluated on two different field datasets collected with above water methodology
288 and two different satellite datasets (HICO and VIIRS).289 **3.1 QWIP applied to field datasets**290 QWIP scores were calculated for two different above water field datasets for evaluation across a
291 variety of water types with different instruments.292 **Quality-Controlled Data.** The QWIP approach was tested on a highly calibrated above-water
293 dataset collected from two different moorings (PANTHYR) (Vansteenkiste et al., 2019;
294 Vanhellemont, 2020). The Ostend dataset ($n=1622$) contained primarily water that clustered as the
295 water type 16 from (Wei et al., 2016) with a strongly green-peaked spectral shape with a
296 fluorescence/red edge characteristic of high phytoplankton. In contrast, the PANTHYR Aqua Alta
297 dataset ($n=4945$) contained primarily water that ranged from Wei et al. (2016) types 6-10 with a high
298 dynamic spectral range, more rounded blue/green spectral shape and very little fluorescence/red edge
299 features. The mean QWIP score was 0.0135 with a small standard deviation of only 0.03 indicating
300 that the data closely followed the mean tendency of the QWIP polynomial (**Figure 5A,B**). Over 98.7
301 % of the data had a QWIP score within ± 0.1 and 100 % within ± 0.2 with a maximum magnitude of
302 $|0.13|$.303 When compared to the spectral quality score proposed by Wei et al. (2016) (“Wei score”), the Wei
304 score for the blue-green data of Aqua Alta were all >0.7 with a mean Wei score of 0.99. However,
305 the Wei score predicted lower values for the more green-peaked water types of Ostend with a mean
306 score of 0.93. Although all spectra passed QWIP, there were 13 spectra deemed failing with Wei
307 scores <0.5 (**Figure 5C**) and 84 spectra of lower quality with Wei Scores of 0.5-0.75 (**Figure 5D**).
308 The relationship between QWIP and Wei Score was not linear and did not follow the negative trend
309 predicted by each method (**Figure 5E**), likely because of the use of only a select number of
310 predefined water types using multi-spectral information in the Wei method, but both methods passed
311 99.8 % of the spectra.312 **Raw WISP-3 Data.** In addition to highly quality controlled data, we assessed how well the QWIP
313 method would identify outliers in a raw dataset that included measurements made with the lens cap
314 on as well as non-water targets (**Figure 6**) ($n=869$). In general, the data across all water types
315 followed the same patterns from the CASCK-P training dataset with QWIP scores within ± 0.2
316 (**Figure 6A,B**). The high QWIP scores were related to the outliers of bad data with unusual spectral
317 shapes or noisy data taken with too little light at dusk (**Figure 6D**). False positives did occur with the
318 QWIP approach where a failing spectrum had a passing QWIP score. The circled region of **Figure**
319 **6A**, for example, indicated the presence of several brown water data points that coincidentally fell
320 within the passing region of the QWIP polynomial. These data had errant spectral shapes that could
321 be easily screened out with some other quality control test such as adding an additional screening for
322 the appropriate range of AVW for a given water type (“blue”, “green”, “brown”) or the Wei score
323 (see below). Specifically, the AVW values of the brown water type was much too low in this instance
324 and could easily be identified with a simple screen of AVW values less than 540 nm (**Figure 6E**).
325 From the training data here, the acceptable ranges in AVW calculated from 400 nm to 700 nm for
326 different water types are: blue-green points ranging 440 nm to 530 nm, green ranging from 510 nm to
327 580 nm and brown water from 550 nm to 590 nm (**Figure 7**).

328 In addition to major outliers, the QWIP score can be useful for diagnosing subtle issues with data
 329 quality. For example, R_{rs} of green outliers with slightly negative scores revealed spectra that were
 330 reasonable overall in spectral shape (**Figure 6C**), but had potentially too low values in the blue range
 331 shorter than 440 nm. This is likely due to a high uncertainty in the removal of surface reflected
 332 skylight and potentially too high of a ρ value. Hence, the QWIP diagnostic can lead to a more
 333 nuanced processing of above water spectra.

334 We compared the absolute value of the QWIP score with the spectral quality score proposed by Wei
 335 et al. (2016) (“Wei score”) on the raw WISP data (**Figure 8**). As tabulated in **Table 1**, there was
 336 consistency between the two approaches using an absolute value $|QWIP|$ threshold of >0.2 and a Wei
 337 score of <0.5 , with 737 data points identified as passing in both techniques and 54 data points as
 338 failing with both techniques (highlighted as the colored portions of **Figure 8**). The 12 data points
 339 with passing Wei scores and failing QWIP scores had QWIP scores that were just slightly above 0.2
 340 and would pass a screen of QWIP threshold of 0.3, however at the expense of passing spectra in the
 341 QWIP score that had failed the Wei score. The Wei approach did correctly fail the “false positive”
 342 brown water spectra discussed above. However, there was discrepancy between the two methods in
 343 roughly 5% of the dataset. Similar to the PANTHYR data, the Wei score was potentially too low for
 344 green data with large red edge reflectance and certain brown water types (**Table 1**). Such waters can
 345 be challenging to assess given the limited band set, particularly in orange and red wavelengths, used
 346 in the Wei method. Barnes et al. (2019) also found that several seemingly high quality spectra had
 347 very low Wei scores.

348 **Table 1.** Comparison of approaches for WISP Lifewatch data

QWIP	Wei Score	
	Pass (>0.50)	Fail (<0.5)
Pass (<0.2)	737 (B=25;G=641;Br=72) ^a	39 ^b (B=0;G=23;Br=16 ^b)
Fail (>0.2)	12 (B=2;G=10;Br=0)	54 (B=8;G=23;Br=23)

349 ^aNumber of datapoints of B=Blue-green, G=Green, and Br=Brown water types

350 ^b10 of these brown data were deemed “false positives” in the QWIP method

351 3.2 QWIP applied to satellite datasets

352 The QWIP procedure was applied to a hyperspectral HICO image (H2012236112610), covering the
 353 complex waters of the Nile Delta, and extending north into the clear waters of the Mediterranean Sea
 354 (**Figure 9A**). Pseudo-true color composites of the HICO imagery can be found in the Supplemental
 355 Materials (**Figure S1**). For this demonstration, a binary exclusion flag was set to distinguish spectra
 356 that pass/fail a nominal threshold value, as defined below. Using rigid exclusion criteria, 99.55 % of
 357 the spectra from the atmospherically-corrected HICO image passed quality control with QWIP score
 358 $< \pm 0.2$, while 0.29 % fell above the QWIP score threshold, and 0.16 % fell below the QWIP score
 359 threshold (**Figure 9B**). The average gradient of integral-normalized $R_{rs}(\lambda)$ spectra (as a function of
 360 AVW) are plotted for these three scenarios, showcasing instances in which data were above/failed
 361 (**Figure 9C**), between/passed (**Figure 9D**), and below/failed (**Figure 9E**) the threshold QWIP score.
 362 Most flagged spectra were characterized by continuous negative reflectance either below 500 nm, or
 363 above 600 nm, and sometimes both. Note, that anomalous negative data should have been screened
 364 out in the SeaDAS processing prior to this analysis.

365 Notably, a series of quality flags already exist within the SeaDAS framework to identify pixels that
 366 fail various levels of quality control (<https://oceancolor.gsfc.nasa.gov/atbd/ocl2flags/>). Pixels
 367 identified as land (LAND flag), radiance saturation (HILT flag), and clouds or ice (CLDICE flag) are
 368 masked during processing (blue in **Figure 10**), thus no R_{rs} data were returned for these pixels. The
 369 only other standard flags that were indicated over water for this scene were the ATMWARN and
 370 MAXAERITER, which indicate a warning in the atmospheric correction procedure, and that the
 371 aerosol model reached the maximum amount of iterations, respectively. The ATMWARN flag
 372 included all instances of MAXAERITER, and this flag comprises <1 % of the ocean pixels (**Figure**
 373 **10**) in the HICO scene. The QWIP identified questionable pixels from regions along the edge of the
 374 scan line, inland, and even a few offshore patches that were not flagged by ATMWARN (brown
 375 regions). In addition, some of the pixels flagged with ATMWARN had spectral shapes with passing
 376 QWIP scores (yellow regions).

377 A series of additional HICO scenes representing a diverse range of optical water types
 378 (H2012237230813, Columbia River outflow, Oregon, USA; H2014191103614, Danube River
 379 Outflow, Romania; H2009344060219, Queensland, Australia; see Figures S1B-D) were sub-sampled
 380 by a range of incremental QWIP score threshold values, to illustrate the connection between QWIP
 381 score and spectral shape. **Figures 11(A-F)** identify more satellite-derived spectra that failed the
 382 nominal quality control criteria, as color coded by discretized QWIP scores. The spectra are often
 383 characterized by sharp increases or decreases in the blue range of $R_{rs}(\lambda)$ and/or contain negative
 384 values. In most cases, as spectral data increasingly deviate from the polynomial relationship between
 385 AVW and NDI(492,665), the anomalous spectral features become more prominent.

386 3.3 QWIP applied to multi-spectral validation data

387 While the QWIP was developed for hyperspectral measurements, the approach can be applied to
 388 multi-spectral data using sensor-specific coefficients to derive the hyperspectral-equivalent AVW as
 389 per Vandermeulen et al. (2020). Here, we test the algorithm efficacy using multi-spectral validation
 390 measurements (in this case, for SNPP-VIIRS) retrieved from NASA's SeaBASS. In order to use the
 391 QWIP as defined in this manuscript, the AVW derived from multi-spectral measurements must first
 392 be translated to a hyperspectral-equivalent AVW through the use of sensor-specific polynomial
 393 offsets (Vandermeulen et al. 2020). The most recent updates to the coefficients for all satellite
 394 sensors processed by OBPG have been developed and published (Vandermeulen, 2022) (also see
 395 Matlab scripts in Supplemental Material).

396 Applying the same procedure as the HICO analysis, 5.2 % and 12.1 % of the VIIRS satellite data
 397 were flagged as falling above/below a QWIP threshold of 0.2, respectively, and 0.2 % and 4.2 % of
 398 *in situ* data were flagged as falling above/below the QWIP threshold of 0.2. Note, given that SNPP-
 399 VIIRS has the fewest number of spectral channels relative to the other validation data streams, it
 400 exhibits a higher uncertainty in the calibrated AVW values relative to many other heritage sensors,
 401 with a mean absolute error (MAE) of 1.21 nm, and bias of -0.15 nm. Such uncertainty is expected
 402 when converting a product derived from 5 bands into an approximation of its hyperspectral
 403 equivalent. To account for this additional uncertainty, we chose here to relax the nominal threshold
 404 value to 0.3 for subsequent reporting, as the 0.2 threshold appeared too stringent for practical
 405 application.

406 Using a nominal threshold of 0.3, we found that 2.1 % and 8.2 % of the VIIRS satellite data were
 407 flagged as falling above/below the QWIP threshold, respectively, and 0.0 % and 1.0 % of *in situ* data
 408 were flagged as falling above/below the QWIP threshold. The flagged spectra were mostly

409 characterized by elevated blue reflectance when the QWIP scores are above the threshold (**Figure**
 410 **12A**), and negative/depressed blue reflectance when the QWIP scores are below the threshold
 411 (**Figure 12B**). A scatter plot comparison between *in situ* (410 nm) and VIIRS (410 nm) R_{rs} matchups
 412 (**Figure 12C**) shows that the flagged outliers tend to form tightly grouped clusters with independent
 413 trends that deviate significantly from the overall linear regression fit. A modest reduction can be seen
 414 in the MAE, and bias between the *in situ* v. satellite matchups for R_{rs} (412) (**Figure 12C**), while
 415 differences at all other wavelengths were negligible. Notably, the removal of flagged pixels improves
 416 the matching of data frequency distributions at R_{rs} (412) (**Figure 12E**) relative to the native dataset
 417 (**Figure 12D**).

418 **4 Discussion, and Future Outlook**

419 The QWIP approach provides a simple quantitative tool to evaluate the quality of both field and
 420 satellite spectra. The QWIP polynomial based on Apparent Visible Wavelength (AVW) and a red and
 421 green band difference index was developed using a broad global training dataset that included blue,
 422 blue-green, green, very green with strong fluorescence, and turbid brown waters. A QWIP score
 423 represents the spectral deviation from the polynomial with scores less than 0.2 generally considered
 424 to be passing. Here, we show how the QWIP can be useful to diagnose major and minor outliers and
 425 potentially correctable spectral anomalies like inaccurate removal of surface reflected skylight from
 426 above water measurements and provide a quantitative means to screen databases for realistic aquatic
 427 spectra both in magnitude and spectral shape. It also has high utility for evaluating different
 428 approaches for atmospheric correction of satellite imagery.

429 The AVW values calculated from field data collected in turbid brown and bright green waters
 430 overlapped in magnitude with each other across red wavelengths and were all less than 600 nm. This
 431 result is different from the initial work of Vandermeulen et al. (2020) who found values of AVW
 432 with increasing chlorophyll up to 617 nm using a synthetic database (Craig et al., 2020). This
 433 suggests caution when using synthetic data that has highly sloped backscattering and other features
 434 (e.g. inelastic scattering) that may not be representative of real world conditions. The brown and
 435 green waters could potentially be better separated with AVW by increasing the spectral range into the
 436 near infrared (NIR) out to 800 nm. Tuning the QWIP with AVW calculated from the ultraviolet to
 437 NIR wavelengths could provide further discrimination of spectral quality and will be further
 438 considered in the future as more datasets become available that cover this larger spectral domain.

439 Spectra with negative QWIP scores resulted from data that were overcorrected for surface reflected
 440 skylight, resulting in lower than expected magnitude in the blue end of the spectrum. Additionally,
 441 several optically shallow spectra that were included in the CASCK training dataset were identified as
 442 having lower than expected QWIP scores due to the sharp increase from red to green wavelengths.
 443 Future research will evaluate how sensitive this metric is to common issues related to surface
 444 reflected skylight, glint, foam, sensor tilt, and other common issues related to collection of field
 445 spectra. Additional analyses will also be conducted to see whether the method can be further adapted
 446 to identify a wide variety of optically shallow water spectra including coral reefs, seagrasses, and
 447 other benthic features (Garcia et al., 2018, 2020).

448 While the QWIP was useful for quality control of a raw field dataset, our research shows that it
 449 cannot be the sole quality flag used to process raw datasets. We found obviously bad spectra (i.e.,
 450 dark spectra with lens covered) that coincidentally had a passing QWIP score. With our training data,
 451 these false positives were easily identified by adding an additional screen to limit the range of
 452 acceptable AVW for blue, green and brown water types (see **Figure 7**) or assessing quality using an

453 addition approach like the spectral-matching approach (Wei et al., 2016). While the two approaches
 454 yielded very similar results, differences were found between QWIP and Wei scores for certain water
 455 types. Specifically, the Wei approach had challenges in assessing the quality of certain green and
 456 brown waters compared to QWIP (see **Figures 5 & 8**).

457 The QWIP approach also proved useful as a quality control metric applied to satellite data. Here,
 458 QWIP was successfully used to flag questionable data from atmospherically-corrected ocean color
 459 satellite imagery from the hyperspectral HICO, as well as the multi-spectral SNPP-VIIRS. Satellite
 460 processing chains do have flags that can identify pixels that fail various levels of quality control;
 461 however, those flags can still let through some questionable spectra. Comparisons between VIIRS
 462 and *in situ* match-up data, for example, showed that removal of pixels with high absolute QWIP
 463 Scores improved the correspondence between the field and satellite data. Additional comparisons
 464 between satellite data which is not masked by quality flags would also prove useful. Further
 465 improvement could involve development and tuning of sensor-specific polynomial offsets used to
 466 extrapolate AVW for multi-channel sensors for more optically complex waters. The technique could
 467 also be tested for 3-4 channel broadband “RGB-type” sensors allowing for greater uncertainty.

468 It is worth emphasizing that the discontinuity in multi-spectral sampling relative to the more
 469 continuous hyperspectral measurements creates a bias in the AVW values, and that the QWIP values
 470 presented in this manuscript are specifically relevant for 1 nm interval hyperspectral R_{rs} data. This
 471 relationship between multi- and hyperspectral AVW is non-linear, and varies as a function of spectral
 472 channels and bandwidth (Vandermeulen et al. 2020, 2022). In order to use the QWIP on multi-
 473 spectral data streams, one of two methods may be employed: 1) convert the multispectral AVW
 474 values to a hyperspectral equivalent value, following Vandermeulen (2022), and then utilize the
 475 QWIP relationship as presented in this manuscript, or 2) an alternative approach is to derive an
 476 independent multispectral QWIP relationship by subsampling a library of quality-controlled hyper-
 477 spectral R_{rs} to the relevant multispectral wavelengths, and use this information to derive a new QWIP
 478 polynomial that would only be used with data of that specific spectral resolution.

479 Because of its simplicity and highly visual component, we foresee high applicability for the QWIP
 480 method across a wide array of applications. Data and scripts we provide in the supplementary
 481 material (see Supplemental Material) would enable users to employ either approach. The quality of
 482 various atmospheric correction routines can be quickly assessed by comparing retrieved results to the
 483 QWIP. The QWIP can also be used to fine tune approaches used to remove surface reflected skylight
 484 and other data processing choices from above water reflectance measurements. The QWIP method
 485 provides a simple tool to help evaluate spectral shape and magnitude for a variety of aquatic water-
 486 leaving reflectance spectra. Indeed, most researchers are well-equipped to apply the QWIP method to
 487 “qwip” their data into shape.

488 5 Conflict of Interest

489 *The authors declare that the research was conducted in the absence of any commercial or financial
 490 relationships that could be construed as a potential conflict of interest.*

491 6 Author Contributions

492 H.D. and R.V. wrote and conceived of this approach and did the primary data analysis and
 493 manuscript preparation. B.B. calculated Wei scores. B.B., A.C., E.K. and Q.V. contributed data,
 494 wrote methods, and contributed to the final editing of the manuscript. Preliminary PANTHYR data
 495 were obtained thanks to platform access, maintenance and installation support of the Institute of

496 Marine Sciences of the Italian National Research Council (CNR-ISMAR) for the AAOT platform,
497 and The Flemish Marine Institute (VLIZ) and POM West-Vlaanderen for the Blue Accelerator
498 platform.

499 **7 Funding**

500 Funding for this research was provided by BELSPO Stereo III TIMBERS project (SR/00/381) and
501 NASA Ocean Biology and Biogeochemistry (80NSSC20K1518). A.C. acknowledges funding
502 provided by BELSPO Stereo III PHYTOBEL project (SR/02/213).

503 **8 Acknowledgments**

504 The authors acknowledge all of the individuals who participated in collection of the data used in this
505 paper.

506 **9 References**

507 Aurin, D. A. (2022). *Hyperspectral In situ Support for PACE*. NASA Goddard Space Flight Center
508 Available at: <https://github.com/nasa/HyperInSPACE>.

509 Bailey, S. W., and Werdell, P. J. (2006). A multi-sensor approach for the on-orbit validation of ocean
510 color satellite data products. *Remote Sensing of Environment* 102, 12–23.

511 Bailey, S. W., Franz, B. A., and Werdell, P. J. (2010). Estimation of near-infrared water-leaving
512 reflectance for satellite ocean color data processing. *Optics express*, 18(7), 7521–7527.

513 Balasubramanian, S. V., Pahlevan, N., Smith, B., Binding, C., Schalles, J., Loisel, H., et al. (2020).
514 Robust algorithm for estimating total suspended solids (TSS) in inland and nearshore coastal waters.
515 *Remote Sensing of Environment*, 111768.

516 Barnes, B. B., Cannizzaro, J. P., English, D. C., and Hu, C. (2019). Validation of VIIRS and MODIS
517 reflectance data in coastal and oceanic waters: An assessment of methods. *Remote Sensing of
518 Environment* 220, 110–123.

519 Bulgarelli, B., and Zibordi, G. (2020). Adjacency radiance around a small island: implications for
520 system vicarious calibrations. *Applied Optics* 59, C63–C69.

521 Casey, K. A., Rousseaux, C. S., Gregg, W. W., Boss, E., Chase, A. P., Craig, S. E., et al. (2020). A
522 global compilation of in situ aquatic high spectral resolution inherent and apparent optical property
523 data for remote sensing applications. *Earth System Science Data* 12, 1123–1139. doi:10.5194/essd-
524 12-1123-2020.

525 Castagna, A., Simis, S., Dierssen, H., Vanhellemont, Q., Sabbe, K., and Vyverman, W. (2020).
526 Extending Landsat 8: Retrieval of an Orange contra-Band for Inland Water Quality Applications.
527 *Remote Sensing* 12, 637. doi:10.3390/rs12040637.

528 Castagna, A., L. Amadei Martínez, M. Bogorad, I. Daveloose, R. Dassevile, H.M. Dierssen, M.
529 Beck, J. Mortelmans, A.I. Dogliotti, D. Doxaran, K. Ruddick, W. Vyverman, K. Sabbe. 2022
530 (submitted). Optical and biogeochemical properties of Belgian inland and coastal waters. *Earth
531 System Science Data*. <https://doi.pangaea.de/10.1594/PANGAEA.940240> (dataset in review)

532 Craig, S. E., Lee, Z., and Du, K. (2020). Top of Atmosphere, Hyperspectral Synthetic Dataset for
 533 PACE (Phytoplankton, Aerosol, and ocean Ecosystem) Ocean Color Algorithm Development.
 534 *National Aeronautics and Space Administration*. doi:10.1594/PANGAEA.915747.

535 Dierssen, H., Bracher, A., Brando, V., Loisel, H., and Ruddick, K. (2020). Data needs for
 536 hyperspectral detection of algal diversity across the globe. *Oceanography* 33, 74–79.

537 Dierssen, H. M., Ackleson, S. G., Joyce, K., Hestir, E., Castagna, A., Lavender, S. J., et al. (2021).
 538 Living up to the Hype of Hyperspectral Aquatic Remote Sensing: Science, Resources and Outlook.
 539 *Frontiers in Environmental Science* 9, 134.

540 Dierssen, H. M., Zimmerman, R. C., Leathers, R. A., Downes, T. V., and Davis, C. O. (2003). Ocean
 541 color remote sensing of seagrass and bathymetry in the Bahamas Banks by high resolution airborne
 542 imagery. *Limnol. Oceanogr.* 48, 444–455.

543 Gao, B. C., and Davis, C. O. (1997). Development of a line-by-line-based atmosphere removal
 544 algorithm for airborne and spaceborne imaging spectrometers. In *Imaging Spectrometry III* (Vol.
 545 3118, pp. 132-141). International Society for Optics and Photonics.

546 Garcia, R. A., Lee, Z., Barnes, B. B., Hu, C., Dierssen, H. M., and Hochberg, E. J. (2020). Benthic
 547 classification and IOP retrievals in shallow water environments using MERIS imagery. *Remote
 548 Sensing of Environment* 249, 112015.

549 Garcia, R. A., Lee, Z., and Hochberg, E. J. (2018). Hyperspectral Shallow-Water Remote Sensing
 550 with an Enhanced Benthic Classifier. *Remote Sensing* 10, 147.

551 Gordon, H. R., & Wang, M. (1994). Retrieval of water-leaving radiance and aerosol optical thickness
 552 over the oceans with SeaWiFS: a preliminary algorithm. *Applied optics*, 33(3), 443-452.

553 Gould, R. W., Arnone, R. A., and Sydor, M. (2001). Absorption, scattering, and remote sensing
 554 reflectance relationships in coastal waters: Testing a new inversion algorithm. *J. Coastal Res.* 17,
 555 328–341.

556 Hommersom, A., Kratzer, S., Laanen, M., Ansko, I., Ligi, M., Bresciani, M., et al. (2012).
 557 Intercomparison in the field between the new WISP-3 and other radiometers (TriOS Ramses, ASD
 558 FieldSpec, and TACCS). *Journal of Applied Remote Sensing* 6, 063615.

559 Ibrahim, A., Franz, B., Ahmad, Z., Healy, R., Knobelispiesse, K., Gao, B. C., et al. (2018).
 560 Atmospheric correction for hyperspectral ocean color retrieval with application to the Hyperspectral
 561 Imager for the Coastal Ocean (HICO). *Remote Sensing of Environment*, 204, 60-75.

562 Knaeps, E., Ruddick, K. G., Doxaran, D., Dogliotti, A. I., Nechad, B., Raymaekers, D., et al. (2015).
 563 A SWIR based algorithm to retrieve total suspended matter in extremely turbid waters. *Remote
 564 Sensing of Environment* 168, 66–79.

565 Lee, Z., Wei, J., Shang, Z., Garcia, R., Dierssen, H. M., Ishizaka, J., et al. (2020). “On-Water
 566 Radiometry Measurements: Skylight-Blocked Approach and Data Processing,” in *IOCCG Ocean
 567 Optics & Biogeochemistry Protocols for Satellite Ocean Colour Sensor Validation*.

568 Mobley, C. D. (1999). Estimation of the remote sensing reflectance from above-surface
 569 measurements. *Applied Optics* 38, 7442–7455.

570 Mortelmans, J., Deneudt, K., Cattrijssse, A., Beauchard, O., Daveloose, I., Vyverman, W., et al.
 571 (2019). Nutrient, pigment, suspended matter and turbidity measurements in the Belgian part of the
 572 North Sea. *Scientific data* 6, 1–8.

573 Ocean Biology Processing Group (2022). *NASA SeaDAS Software Package*. Available at:
 574 <https://seadas.gsfc.nasa.gov>.

575 O'Reilly, J. E., Maritorena, S., Mitchell, B. G., and Siegel, D. A. (1998). Ocean color chlorophyll
 576 algorithms for SeaWiFS. *Journal of Geophysical Research* 103, 24,937-24,953.

577 Qin, P., Simis, S. G., and Tilstone, G. H. (2017). Radiometric validation of atmospheric correction
 578 for MERIS in the Baltic Sea based on continuous observations from ships and AERONET-OC.
 579 *Remote sensing of environment* 200, 263–280.

580 Röttgers, R., Doerffer, R., McKee, D., and Schönfeld, W. (2016). *The Water Optical Properties
 581 Processor (WOPP): Pure Water Spectral Absorption, Scattering, and Real Part of Refractive Index
 582 Model*. ATBD.

583 Ruddick, K. G., Voss, K., Boss, E., Castagna, A., Frouin, R., Gilerson, A., et al. (2019). A review of
 584 protocols for fiducial reference measurements of water-leaving radiance for validation of satellite
 585 remote-sensing data over water. *Remote Sensing* 11, 2198.

586 Ruddick, Kevin G., De Cauwer, Vera, Park, Young-Je, Moore, Gerald, (2006), Seaborne
 587 measurements of near infrared water-leaving reflectance: The similarity spectrum for turbid waters,
 588 *Limnology and Oceanography*, 51, doi: 10.4319/lo.2006.51.2.1167.

589 Shang, Z., Lee, Z., Wei, J., and Lin, G. (2020). Impact of ship on radiometric measurements in the
 590 field: a reappraisal via Monte Carlo simulations. *Opt. Express, OE* 28, 1439–1455.
 591 doi:10.1364/OE.28.001439.

592 Tilstone, G., Dall'Olmo, G., Hieronymi, M., Ruddick, K., Beck, M., Ligi, M., et al. (2020). Field
 593 intercomparison of radiometer measurements for ocean colour validation. *Remote Sensing* 12, 1587.

594 Tzortziou, M., Herman, J. R., Gallegos, C. L., Neale, P. J., Subramaniam, A., Harding Jr, L. W., et al.
 595 (2006). Bio-optics of the Chesapeake Bay from measurements and radiative transfer closure.
 596 *Estuarine, Coastal and Shelf Science* 68, 348–362.

597 Vandermeulen, R. A. (2022). Apparent Visible Wavelength (AVW): NASA Algorithm Theoretical
 598 Basis Document. Available at: <https://oceancolor.gsfc.nasa.gov/atbd/avw/>.

599 Vandermeulen, R. A., Mannino, A., Craig, S. E., and Werdell, P. J. (2020). 150 shades of green:
 600 Using the full spectrum of remote sensing reflectance to elucidate color shifts in the ocean. *Remote
 601 Sensing of Environment* 247, 111900.

602 Vanhellemont, Q. (2020). Sensitivity analysis of the dark spectrum fitting atmospheric correction for
 603 metre-and decametre-scale satellite imagery using autonomous hyperspectral radiometry. *Optics
 604 Express* 28, 29948–29965.

605 Vansteenwegen, D., Ruddick, K., Cattrijssse, A., Vanhellemont, Q., and Beck, M. (2019). The pan-
606 and-tilt hyperspectral radiometer system (PANTHYR) for autonomous satellite validation
607 measurements—Prototype design and testing. *Remote Sensing* 11, 1360.

608 Voss, K. J., Johnson, C. B., Yarbrough, M. A., Gleason, A., Flora, S. J., Feinholz, M. E., et al.
609 (2017). An overview of the Marine Optical Buoy (MOBY): Past, present and future. *Proceedings of*
610 *the D-240 FRM4SOC-PROC1 Proceedings of WKP-1 (PROC-1) Fiducial Reference Measurements*
611 *for Satellite Ocean Colour (FRM4SOC), Tartu, Estonia*, 8–13.

612 Wei, J., Lee, Z., and Shang, S. (2016). A system to measure the data quality of spectral remote-
613 sensing reflectance of aquatic environments. *Journal of Geophysical Research: Oceans* 121, 8189–
614 8207.

615 Zaneveld, J. R. V. (1994). “Optical closure: from theory to measurement,” in *Ocean Optics* (Oxford
616 University Press).

617 Zhang, X., He, S., Shabani, A., Zhai, P.-W., and Du, K. (2017). Spectral sea surface reflectance of
618 skylight. *Optics Express* 25, A1–A13.

619 Zibordi, G., Mélin, F., Berthon, J.-F., Holben, B., Slutsker, I., Giles, D., et al. (2009). AERONET-
620 OC: a network for the validation of ocean color primary products. *Journal of Atmospheric and*
621 *Oceanic Technology* 26, 1634–1651.

622 Zibordi, G., Voss, K., Johnson, B. C., and Mueller, J. L. (2019). Protocols for satellite ocean color
623 data validation: In situ optical radiometry. *IOCCG Protocols Document*.

624

625

626 **Figure Captions**

627 **Figure 1.** Sampling locations from the global hyperspectral training data (CASCK-P) used to
 628 develop the polynomial spanned open ocean, coastal, and inland waters. Adapted from Casey et al.
 629 (2020).

630 **Figure 2.** A simple screening approach modified from Balasubramanian et al. (2020) was used to
 631 evaluate the QWIP approach for three water types: Blue-green, Green, and Brown. The remote
 632 sensing spectra shown are from the (CASCK-P) training dataset and are identified with color and
 633 shape in Figs. 3 and 4 based on this schema.

634 **Figure 3.** Examples of different spectral band math approaches compared to Apparent Visible
 635 Wavelength (AVW) evaluated with the CASCK-P training dataset separated by water type: Blue-
 636 green (blue circles), Green (green diamonds) and Brown (red squares). These approaches were not
 637 selected for use do to overall fit and divergence of data from different water types.
 638

639 **Figure 4. (A)** The QWIP relationship between Apparent Visible Wavelength (AVW) and the
 640 Normalized Difference Index (NDI) with the CASCK-P training dataset showing the final tuned
 641 QWIP polynomial (thick magenta line) with different levels of QWIP scores (± 0.1 dotted magenta
 642 and ± 0.2 dashed magenta). Water types include: Blue-green (blue circles), Green (green diamonds)
 643 and Brown (red squares). **(B)** Histogram of the QWIP scores from Panel A are predominantly within
 644 ± 0.1 for the training data. **(C)** The remote sensing reflectance (R_{rs}) of outliers with negative QWIP
 645 scores < -0.2 were associated with optically shallow water features. **(D)** Outliers with QWIP scores
 646 > 0.2 exhibited higher blue associated with surface reflected skylight and higher overall magnitude
 647 spectra.
 648

649 **Figure 5. (A)** The QWIP scores from highly quality controlled hyperspectral PANTHYR reflectance
 650 data from Vanhellemont (2020). **(B)** QWIP scores were predominantly within $< \pm 0.1$. Water types
 651 include: Blue-green (blue circles), Green (green diamonds) and Brown (red squares). Remote sensing
 652 reflectance of data with Wei scores **(C)** less than 0.5 **(D)** between 0.5 and 0.75. **(E)** Comparison of
 653 Wei scores (Wei et al. 2016) and the absolute value of QWIP scores for the entire dataset.
 654

655 **Figure 6. (A)** The QWIP approach was used for quality control of a raw WISP dataset. Water types
 656 include: Blue-green (blue circles), Green (green diamonds) and Brown (red squares) following from
 657 Fig. 2. The red circle highlights false positive data of brown water type that coincidentally fall within
 658 the polynomial limits. **(B)** The majority of the data had QWIP scores of ± 0.2 . **(C)** Remote sensing
 659 reflectance (R_{rs}) of green outliers with slightly negative scores had good spectral shapes but too low
 660 in the blue. **(D)** High QWIP scores were related to the outliers of bad data with unusual spectral
 661 shapes. **(E)** Brown outliers with failing spectral shapes were identified as having lower AVW than
 662 expected for the water type (AVW < 540 nm).
 663

664 **Figure 7.** Histogram of the distribution of Apparent Visible Wavelength (AVW) from Blue-green
 665 (blue), Green (green), and Brown (red) water types showing the overlap and general ranges expected
 666 for each water type. Ranges from the CASCK-P training data from Figure 4.
 667

668 **Figure 8.** Comparisons of the absolute value of the QWIP score with the spectral quality score
 669 proposed by Wei et al. (2016) (“Wei score”). A QWIP threshold of > 0.2 and a Wei score of < 0.5 were
 670 considered failing spectra and vice versa. Colored boxes highlight where both approaches pass

671 (green) and fail (yellow) data. Water types include: Blue-green (blue circles), Green (green
 672 diamonds) and Brown (red squares) following from Fig. 2.
 673

674 **Figure 9:** (A) Mapped HICO scene on which the QWIP procedure was tested. (B) The AVW is
 675 compared to NDI, and (C,E) spectra deviating from the QWIP are nominally deemed to fail quality
 676 control criteria, and those (D) within the uncertainty bounds of the polynomial pass. The spectral
 677 color scheme relates to the corresponding AVW values, as defined in (A).

678 **Figure 10:** A binary quality control (QC) map of a HICO satellite image, illustrating locations in
 679 which the NASA processing “l2gen” flags identified pixels with suspect quality (ATMWARN) and
 680 pixels identified as either passing ($|QWIP|<0.2$) or failing ($|QWIP|>0.2$) QC with the QWIP method.
 681 Orange pixels were flagged by both QWIP and l2gen.

682 **Figure 11:** HICO spectra that fail quality control criteria by falling above (A, C, E) or below (B, D,
 683 F) a nominal QWIP threshold (0.2), for a diverse range of images from the (A,B) Columbia River
 684 outflow, USA, (D, E) Danube River outflow, Romania, and (F, G) Queensland, Australia.

685 **Figure 12:** Remote sensing reflectance (R_{rs}) estimated from SNPP-VIIRS that (A) exceeded and (B)
 686 fell below a nominal QWIP score of 0.3. (C) Scatter plot of $R_{rs}(410)$ for *in situ* obtained from the
 687 SeaBASS archive compared to matchup data retrieved from VIIRS imagery. Blue dots represent
 688 passing data with QWIP scores less than ± 0.3 . A modest reduction in mean absolute error and mean
 689 bias between *in situ* and VIIRS measurements was found when only passing values (blue dots) are
 690 used. (D, E) The frequency distribution of data improves after the removal of those spectra flagged
 691 by the QWIP approach.

692

Figure 1.TIF

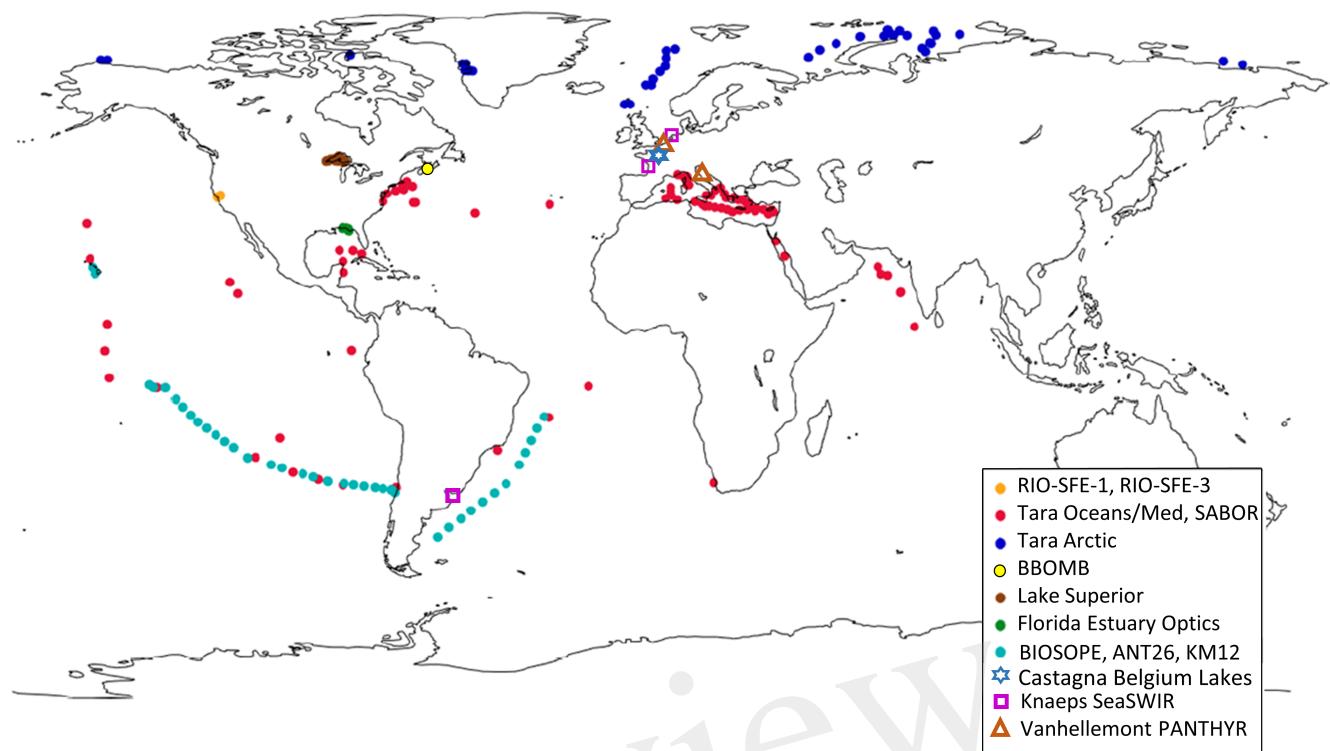


Figure 1. Sampling locations from the global hyperspectral training data (CASCK-P) used to develop the polynomial spanned open ocean, coastal, and inland waters. Adapted from Casey et al. (2020).

Figure 2.TIF

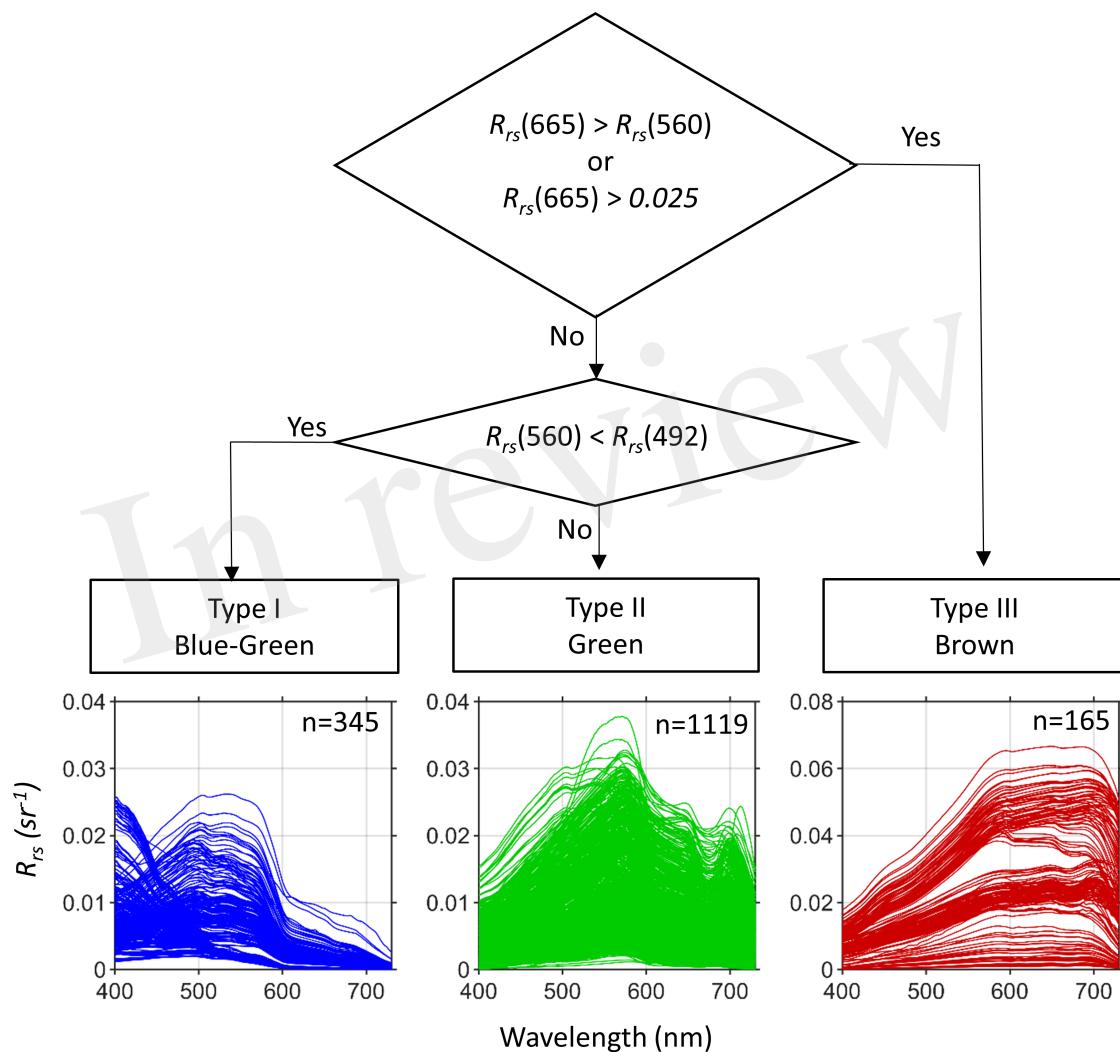


Figure 2. A simple screening approach modified from Balasubramanian et al. (2020) was used to evaluate the QWIP approach for three water types: Blue-green, Green, and Brown. The remote sensing spectra shown are from the (CASCK-P) training dataset and are identified with color and shape in Figs. 3 and 4 based on this schema.

Figure 3.TIF

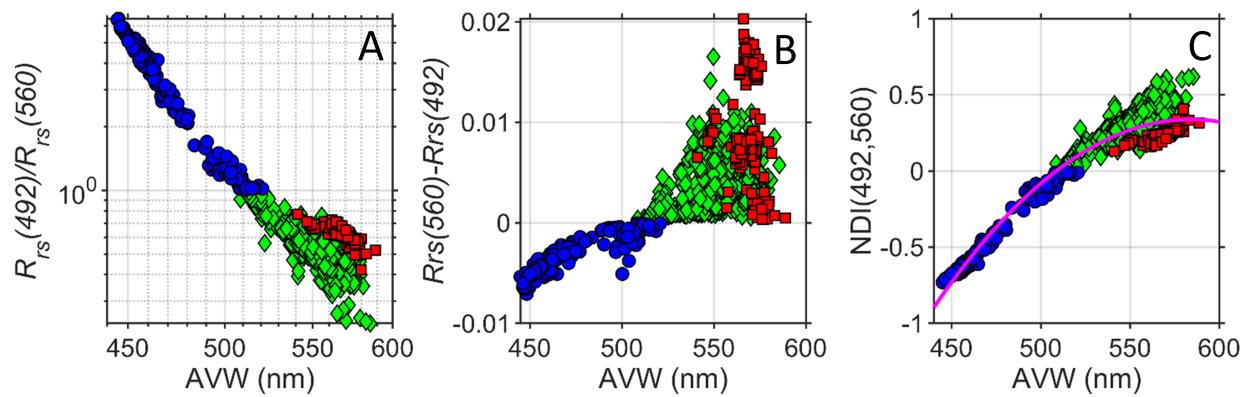


Figure 3. Examples of different spectral band math approaches compared to Apparent Visible Wavelength (AVW) evaluated with the CASCK-P training dataset separated by water type: Blue-green (blue circles), Green (green diamonds) and Brown (red squares). These approaches were not selected for use do to overall fit and divergence of data from different water types.

Figure 4.TIF

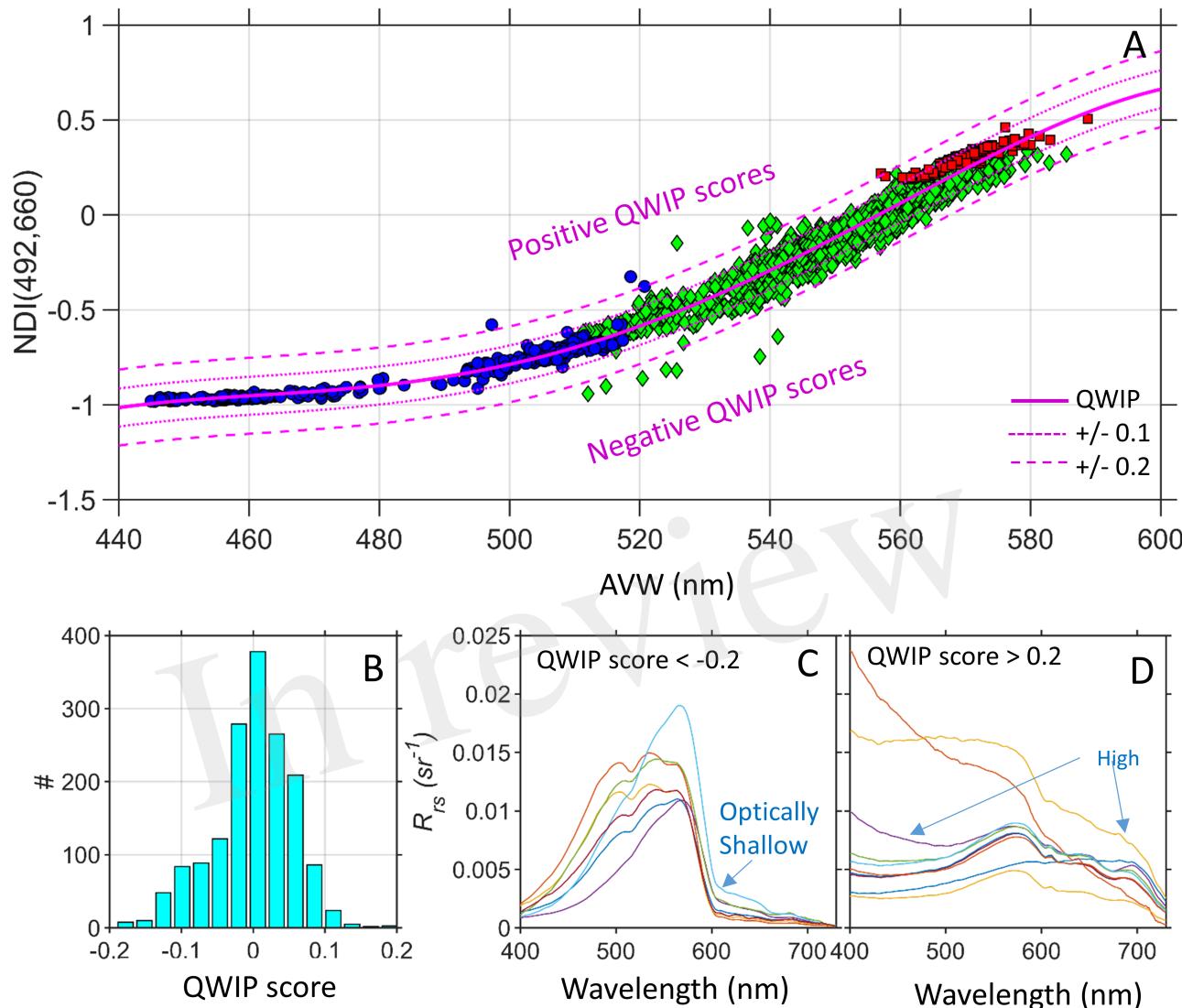


Figure 4. (A) The QWIP relationship between Apparent Visible Wavelength (AVW) and the Normalized Difference Index (NDI) with the CASCK-P training dataset showing the final tuned QWIP polynomial (thick magenta line) with different levels of QWIP scores (± 0.1 dotted magenta and ± 0.2 dashed magenta). Water types include: Blue-green (blue circles), Green (green diamonds) and Brown (red squares). **(B)** Histogram of the QWIP scores from Panel A are predominantly within ± 0.1 for the training data. **(C)** The remote sensing reflectance (R_{rs}) of outliers with negative QWIP scores < -0.2 were associated with optically shallow water features. **(D)** Outliers with QWIP scores > 0.2 exhibited higher blue associated with surface reflected skylight and higher overall magnitude spectra.

Figure 5.TIF

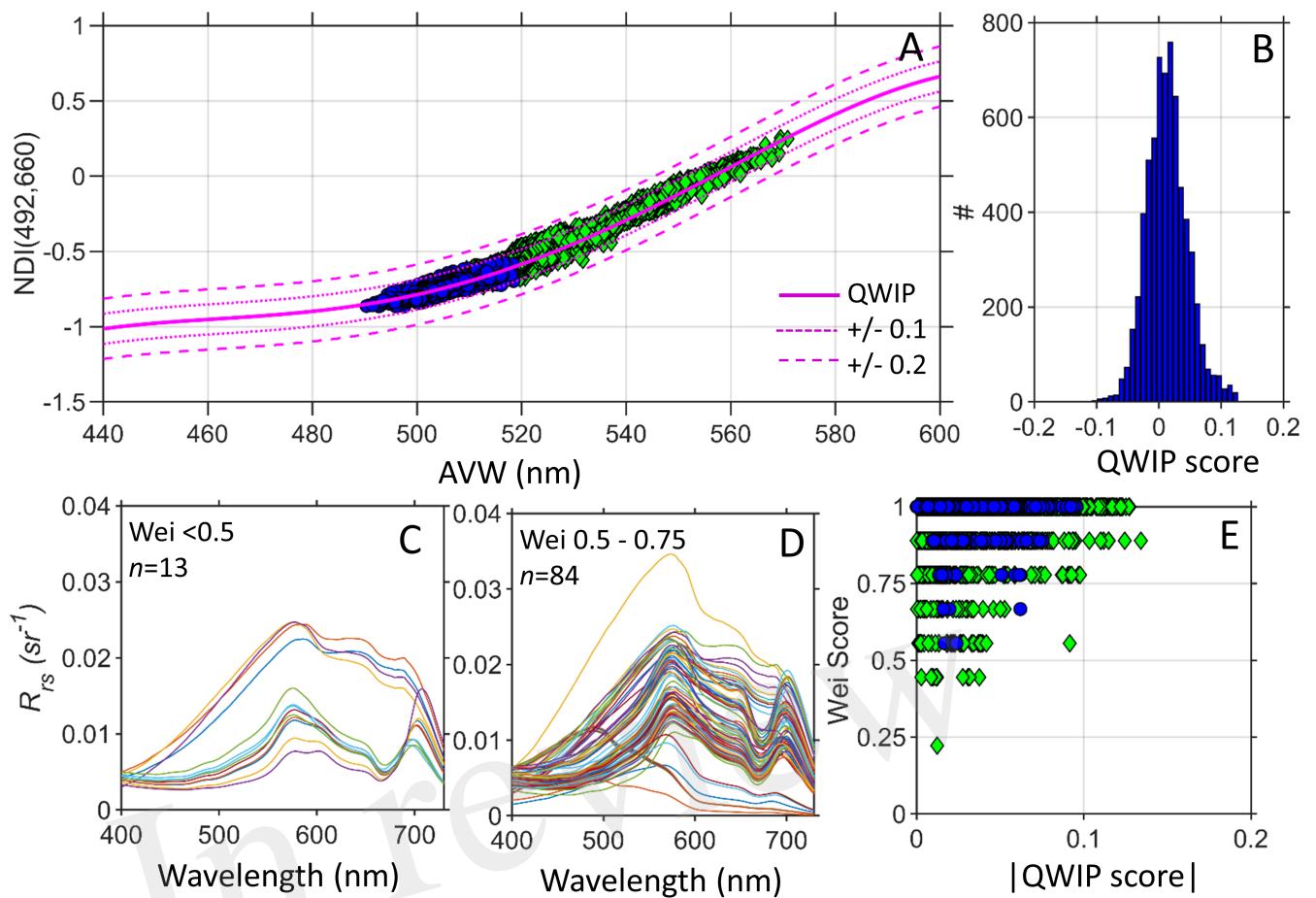


Figure 5. (A) The QWIP scores from highly quality controlled hyperspectral PANTHYR reflectance data from Vanhellemont (2020). (B) QWIP scores were predominantly within $<\pm 0.1$. Water types include: Blue-green (blue circles), Green (green diamonds) and Brown (red squares). Remote sensing reflectance of data with Wei scores (C) less than 0.5 (D) between 0.5 and 0.75. (E) Comparison of Wei scores (Wei et al. 2016) and the absolute value of QWIP scores for the entire dataset.

Figure 6.TIF

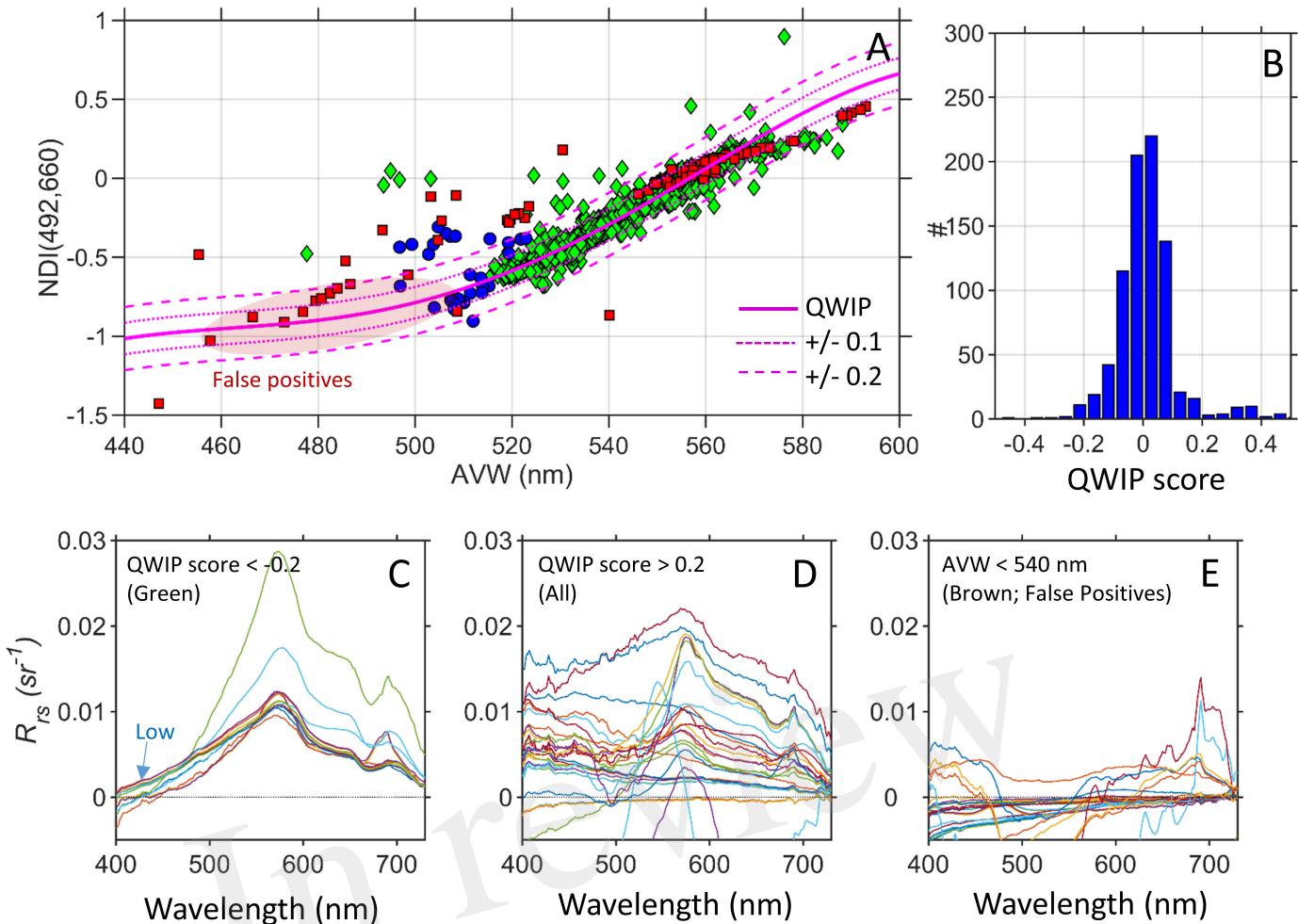


Figure 6. **A)** The QWIP approach was used for quality control of a raw WISP dataset. Water types include: Blue-green (blue circles), Green (green diamonds) and Brown (red squares) following from Fig. 2. The red circle highlights false positive data of brown water type that coincidentally fall within the polynomial limits. **B)** The majority of the data had QWIP scores of ± 0.2 . **C)** Remote sensing reflectance (R_{rs}) of green outliers with slightly negative scores had good spectral shapes but too low in the blue. **D)** High QWIP scores were related to the outliers of bad data with unusual spectral shapes. **E)** Brown outliers with failing spectral shapes were identified as having lower AVW than expected for the water type ($AVW < 540$ nm).

Figure 7.TIF

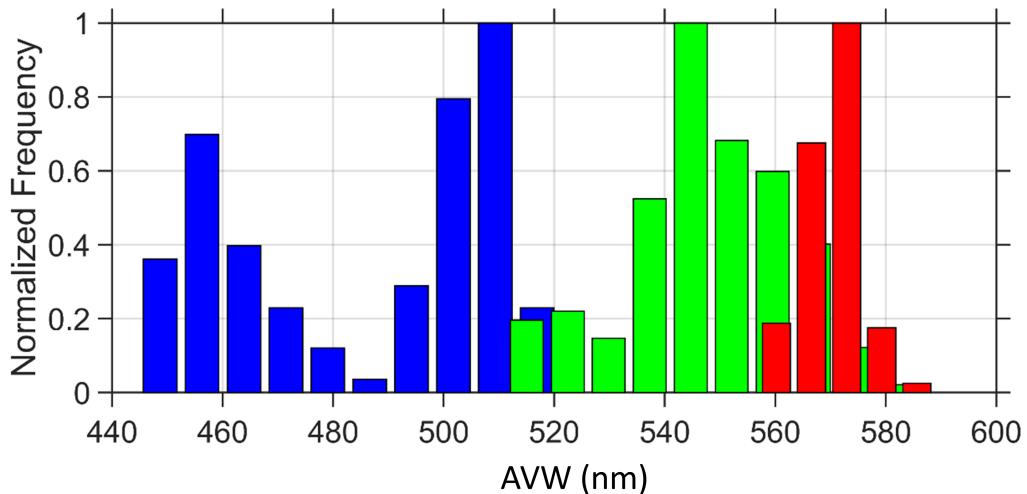


Figure 7. Histogram of the distribution of Apparent Visible Wavelength (AVW) from Blue-green (blue), Green (green), and Brown (red) water types showing the overlap and general ranges expected for each water type. Ranges from the CASCK-P training data from Fig. 4.

Figure 8.TIF

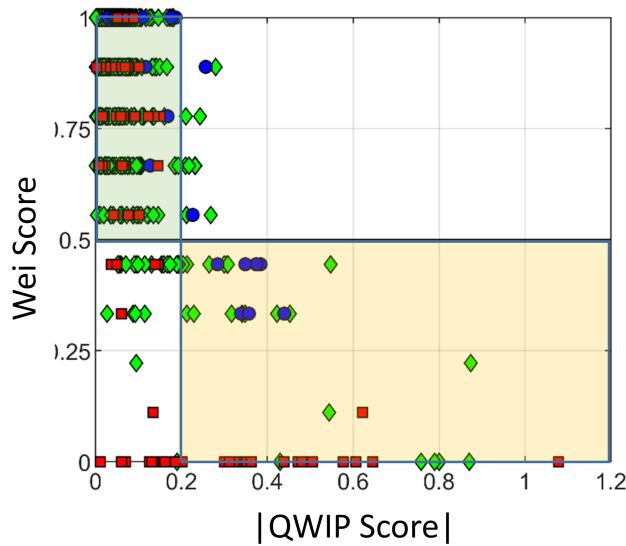


Figure 8. Comparisons of the absolute value of the QWIP score with the spectral quality score proposed by Wei et al. (2016) (“Wei score”). A QWIP threshold of >0.2 and a Wei score of <0.5 were considered failing spectra and vice versa. Colored boxes highlight where both approaches pass (green) and fail (yellow) data. Water types include: Blue-green (blue circles), Green (green diamonds) and Brown (red squares) following from Fig. 2.

Figure 9.TIF

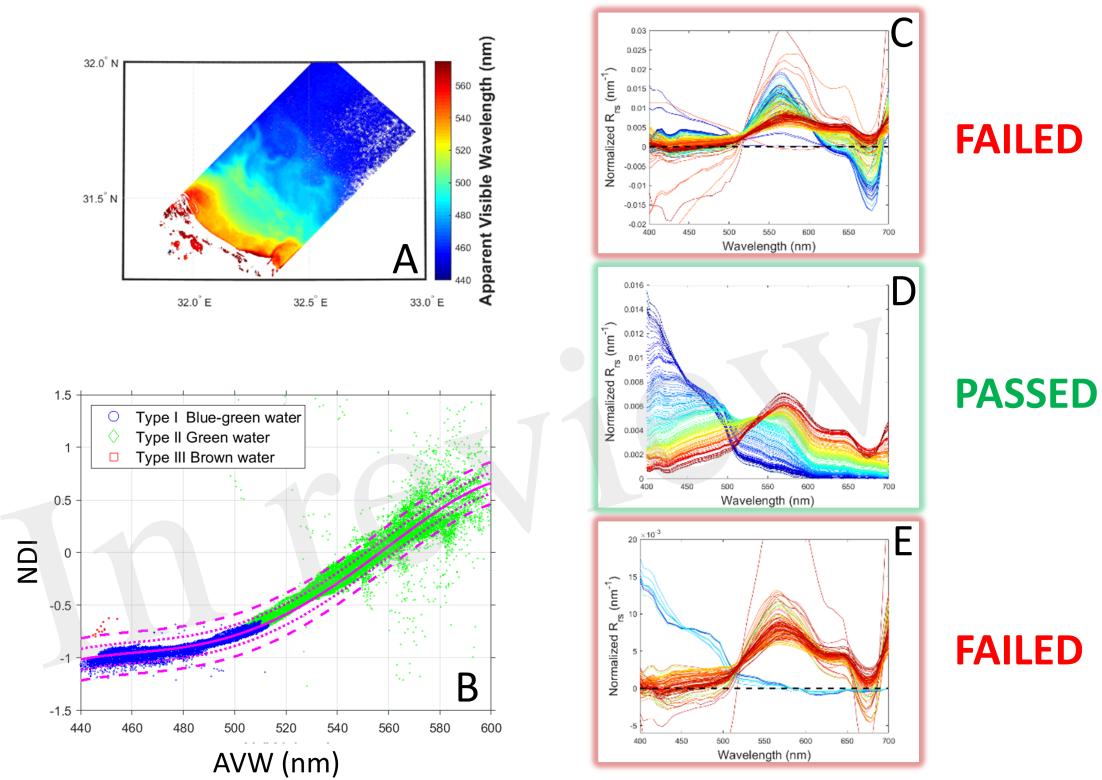


Figure 9: (A) Mapped HICO scene on which the QWIP procedure was tested. (B) The AVW is compared to NDI , and (C,E) spectra deviating from the QWIP are nominally deemed to fail quality control criteria, and those (D) within the uncertainty bounds of the polynomial pass. The spectral color scheme relates to the corresponding AVW values, as defined in (A).

Figure 10.TIF

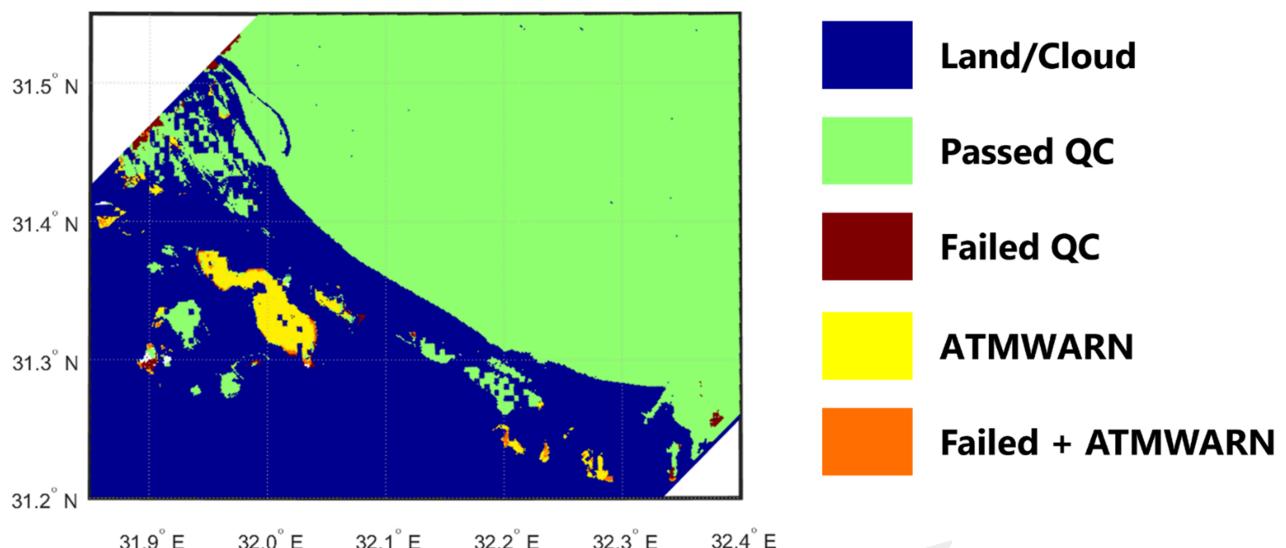


Figure 10: A binary quality control (QC) map of a HICO satellite image, illustrating locations in which the NASA processing “l2gen” flags identified pixels with suspect quality (ATMWARN) and pixels identified as either passing ($|QWIP|<0.2$) or failing ($|QWIP|>0.2$) QC with the QWIP method. Orange pixels were flagged by both QWIP and l2gen.

Figure 11.TIF

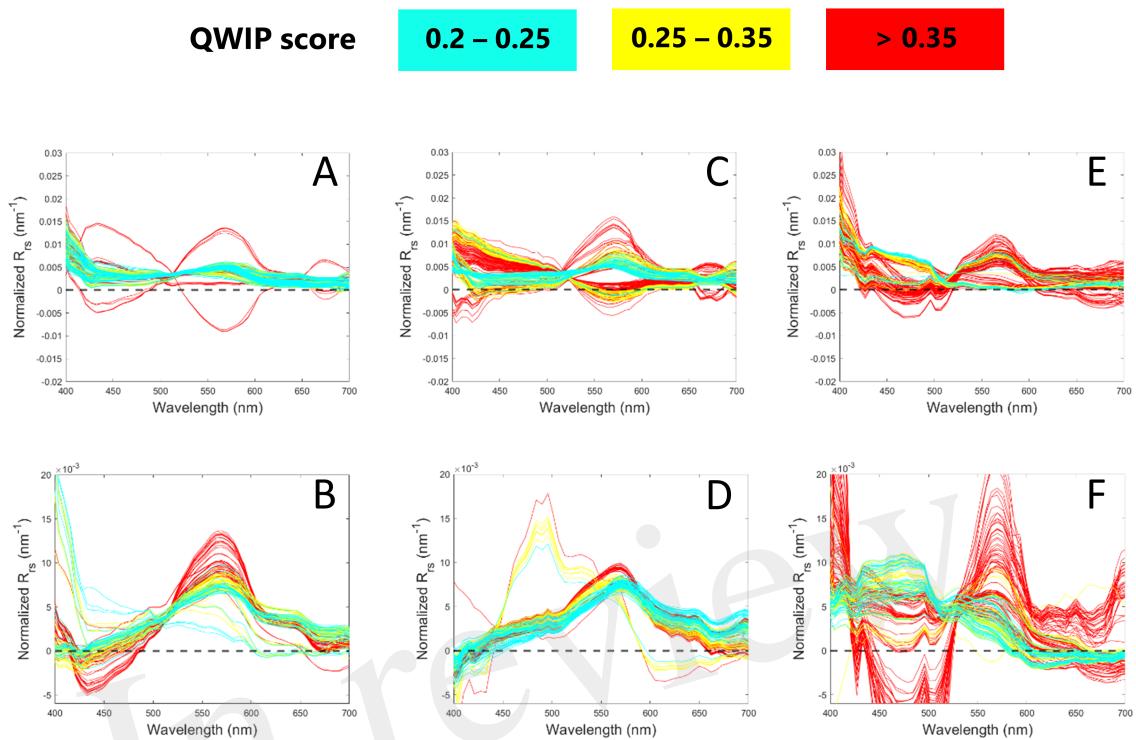


Figure 11: HICO spectra that fail QC criteria by falling above (**A, C, E**) or below (**B, D, F**) a nominal QWIP threshold (0.2), for a diverse range of images from the (**A,B**) Columbia River outflow, USA, (**D, E**) Danube River outflow, Romania, and (**F, G**) Queensland, Australia.

Figure 12.TIF

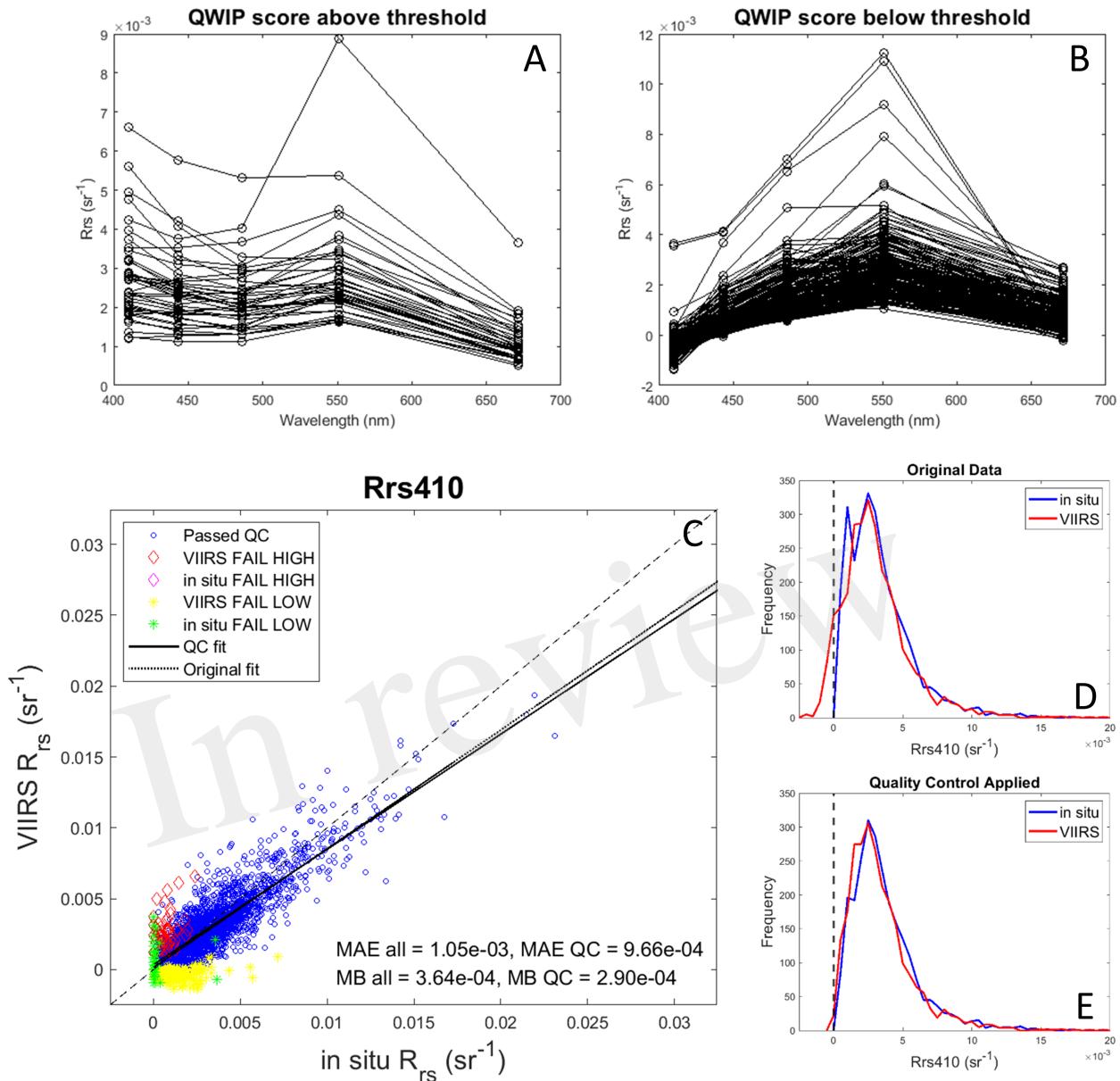


Figure 12: Remote sensing reflectance (R_{rs}) estimated from SNPP-VIIRS that (A) exceeded and (B) fell below a nominal QWIP score of 0.3. (C) Scatter plot of $R_{rs}(410)$ for *in situ* obtained from the SeaBASS archive compared to matchup data retrieved from VIIRS imagery. Blue dots represent passing data with QWIP scores less than ± 0.3 . A modest reduction in mean absolute error and mean bias between *in situ* and VIIRS measurements was found when only passing values (blue dots) are used. (D, E) The frequency distribution of data improves after the removal of those spectra flagged by the QWIP approach.